

sigmod2013-slides

Outline

1. Motivation and Overview
2. Taxonomic Knowledge - Entities and Classes
 1. Scope & Goal
 - Knowledge Bases are labeled graphs
 - Goal: finding classes and instance
 - WordNet is a lexical knowledge base
 - few instances
 - only common nouns as classes
 - only English labels
 2. Wikipedia-centric Methods
 - Wikipedia is a rich source of instances
 - YAGO = WordNet + Wikipedia
 3. Web-based Methods
 - Goal: find instances of classes
 - extract instances
 - build and refine a taxonomy
3. Factual Knowledge - Relations between Entities
 1. Scope & Goal
 1. IE sources
 1. Semi-structured data
 2. Free text
 3. Web tables
 2. IE type
 1. Source-centric IE
 2. Yield-centric IE
 3. Goal: Yield-centric IE
 2. RegEx-based Extraction
 3. Pattern-based Harvesting
 4. Consistency Reasoning
 5. Probabilistic Methods
 6. Web-Table Methods
4. Emerging Knowledge - New Entities & Relations
 1. Open Information Extraction
 2. Relation Paraphrases
 3. Big Data Algorithms
5. Temporal Knowledge - Validity Times of Facts
6. Contextual Knowledge - Entity Name Disambiguation
 1. NERD Problem
 1. named-entity recognition (NER)
 2. co-reference resolution
 3. named-entity disambiguation (NED)
 2. NED Principles
 3. Coherence-based Methods
 4. Rare & Emerging Entities
7. Linked Knowledge - Entity Matching
 1. Goal: Find equivalence classes of entities, and of records
8. Wrap-up

2-5: Big Data Methods for Knowledge Harvesting

6-7: Knowledge for Big Data Analytics

Spectrum of Machine Knowledge

- factual knowledge
- taxonomic knowledge (ontology)
- lexical knowledge (terminology)
- contextual knowledge (entity occurrences, entity-name disambiguation)
- linked knowledge (entity equivalence, entity resolution)
- multi-lingual knowledge
- temporal knowledge (fluents)
- spatial knowledge

- ephemeral knowledge (dynamic services)
- common-sense knowledge (properties)
- common-sense knowledge (rules)
- emerging knowledge (open IE)
- multimodal knowledge (photos, videos)
- social knowledge (opinions)
- epistemic knowledge((un-)trusted beliefs)